



1600

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/581,976B

DATE: 11/26/2002

TIME: 09:26:10

Input Set : A:\seqlist.txt

Output Set: N:\CRF4\11262002\I581976B.raw

```
4 <110> APPLICANT: Dalemans, Wilfried L.J.
        Gerard, Catherine Marie Ghislaine
 7 <120> TITLE OF INVENTION: Vaccine
10 <130> FILE REFERENCE: B45124
                                                         ENTERED
12 <140> CURRENT APPLICATION NUMBER: 09/581,976B
13 <141> CURRENT FILING DATE: 2000-06-20
15 <150> PRIOR APPLICATION NUMBER: PCT/EP98/08563
16 <151> PRIOR FILING DATE: 1998-12-18
18 <150> PRIOR APPLICATION NUMBER: GB 9727262.9
19 <151> PRIOR FILING DATE: 1997-12-24
21 <160> NUMBER OF SEQ ID NOS: 28
23 <170> SOFTWARE: FastSEQ for Windows Version 3.0
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 220
27 <212> TYPE: PRT
28 <213> ORGANISM: Artificial Sequence
30 <220> FEATURE:
31 <223> OTHER INFORMATION: Chimaeric protein (protein D from Haemoplilus
         influenza B and E7 from Human papilloma virus type
33
35 <400> SEQUENCE: 1
36 Met Asp Pro Ser Ser His Ser Ser Asn Met Ala Asn Thr Gln Met Lys
37
                    5
                                        10
38
   Ser Asp Lys Ile Ile Ile Ala His Arg Gly Ala Ser Gly Tyr Leu Pro
39
                                    25
                                                        30
   Glu His Thr Leu Glu Ser Lys Ala Leu Ala Phe Ala Gln Gln Ala Asp
41
42
   Tyr Leu Glu Gln Asp Leu Ala Met Thr Lys Asp Gly Arg Leu Val Val
43
44
   Ile His Asp His Phe Leu Asp Gly Leu Thr Asp Val Ala Lys Lys Phe
45
                                            75
   Pro His Arg His Arg Lys Asp Gly Arg Tyr Tyr Val Ile Asp Phe Thr
46
47
   Leu Lys Glu Ile Gln Ser Leu Glu Met Thr Glu Asn Phe Glu Thr Met
49
               100
                                    105
50
   Ala Met His Gly Asp Thr Pro Thr Leu His Glu Tyr Met Leu Asp Leu
           115
                               120
   Gln Pro Glu Thr Thr Asp Leu Tyr Cys Tyr Glu Gln Leu Asn Asp Ser
                           135
                                                140
54 Ser Glu Glu Glu Asp Glu Ile Asp Gly Pro Ala Gly Gln Ala Glu Pro
                       150
                                           155
   Asp Arg Ala His Tyr Asn Ile Val Thr Phe Cys Cys Lys Cys Asp Ser
56
```

170

165

57

RAW SEQUENCE LISTING DATE: 11/26/2002 PATENT APPLICATION: US/09/581,976B TIME: 09:26:10

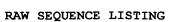
Input Set : A:\seqlist.txt

```
Thr Leu Arg Leu Cys Val Gln Ser Thr His Val Asp Ile Arg Thr Leu
58
59
                180
                                    185
60
   Glu Asp Leu Leu Met Gly Thr Leu Gly Ile Val Cys Pro Ile Cys Ser
61
            195
                                200
62
   Gln Lys Pro Thr Ser Gly His His His His His
63
        210
                            215
                                                220
65 <210> SEQ ID NO: 2
66 <211> LENGTH: 663
67 <212> TYPE: DNA
68 <213> ORGANISM: Artificial Sequence
70 <220> FEATURE:
71 <223> OTHER INFORMATION: Chimaeric protein (protein D from Haemoplilus
72
         influenza B and E7 from Human papilloma virus type
73
75 <400> SEQUENCE: 2
                                                                            60
   atggatccaa gcagccattc atcaaatatg gcgaataccc aaatgaaatc agacaaaatc
    attattqctc accqtqqtqc taqcqqttat ttaccaqaqc atacqttaqa atctaaaqca
                                                                           120
   cttgcgtttg cacaacaggc tgattattta gagcaagatt tagcaatgac taaggatggt
                                                                           180
79
   cgtttagtgg ttattcacga tcacttttta gatggcttga ctgatgttgc gaaaaaattc
                                                                           240
   ccacategte ategtaaaga tggcegttae tatgteateg aetttaeett aaaagaaatt
                                                                           300
   caaaqtttaq aaatqacaqa aaactttqaa accatqqcca tqcatqqaqa tacacctaca
                                                                           360
   ttgcatgaat atatgttaga tttgcaacca gagacaactg atctctactg ttatgagcaa
                                                                           420
83 ttaaatgaca gctcagagga ggaggatgaa atagatggtc cagctggaca agcagaaccg
                                                                           480
   gacagagccc attacaatat tgtaaccttt tgttgcaagt gtgactctac gcttcggttg
                                                                           540
   tgcgtacaaa gcacacacgt agacattcgt actttggaag acctgttaat gggcacacta
                                                                           600
   ggaattgtgt gccccatctg ttctcagaaa ccaactagtg gccaccatca ccatcaccat
                                                                           660
87 taa
                                                                           663
89 <210> SEQ ID NO: 3
90 <211> LENGTH: 822
91 <212> TYPE: DNA
92 <213> ORGANISM: Artificial Sequence
94 <220> FEATURE:
95 <223> OTHER INFORMATION: Chimaeric protein (protein D from Haemoplilus
96
         influenza B and E6 from Human papilloma virus type
97
         16)
99 <400> SEQUENCE: 3
100 atggatecaa geagecatte ateaaatatg gegaatacee aaatgaaate agacaaaate
101 attattgctc accgtggtgc tagcggttat ttaccagagc atacgttaga atctaaagca 120
102 cttqcqtttq cacaacaqqc tqattattta qaqcaaqatt taqcaatqac taaqqatqqt 180
103 cgtttagtgg ttattcacga tcacttttta gatggcttga ctgatgttgc gaaaaaattc 240
104 ccacatcgtc atcgtaaaga tggccgttac tatgtcatcg actttacctt aaaagaaatt 300
105 caaagtttag aaatgacaga aaactttgaa accatggcca tgtttcagga cccacaggag 360
106 cgacccagaa agttaccaca gttatgcaca gagctgcaaa caactataca tgatataata 420
107 ttagaatgtg tgtactgcaa gcaacagtta ctgcgacgtg aggtatatga ctttgctttt 480
108 cgggatttat gcatagtata tagagatggg aatccatatg ctgtatgtga taaatgttta 540
109 aagttttatt ctaaaattag tgagtataga cattattgtt atagtttgta tggaacaaca 600
110 ttagaacagc aatacaacaa accgttgtgt gatttgttaa ttaggtgtat taactgtcaa 660
111 aagccactgt gtcctgaaga aaagcaaaga catctggaca aaaagcaaag attccataat 720
112 ataaggggtc ggtggaccgg tcgatgtatg tcttgttgca gatcatcaag aacacgtaga 780
```



Input Set : A:\seqlist.txt

112	13 gaaacccagc tgactagtgg ccaccatcac catcaccatt aa														· 0	22	
	gaaacccage tgactagtgg ccaccatcac catcaccatt aa 822 < <210> SEQ ID NO: 4														22		
	<211> LENGTH: 273																
	<211> LENGIR: 273 <212> TYPE: PRT																
	<213> ORGANISM: Artificial Sequence																
					ALCI.	LICI	3T 26	eque	ice								
	<220				~ N 4 70 CTT :	T ())]	Ob. I					4			**	, .	, .
	<223.														om Ha	emopli	Tus
123				nza i	s and	ı Eo	ITOI	n Hui	nan 1	papı.	lloma	1 V1.	cus 1	суре			
124	< 4 O O :	16		ο υ .	4												
127	<400:		_			n; o	C ~ ~	C~~	7000	Mot	71.	7.00	mb~	C1-	Mot	T	
128	ме с 1	ASP	PIO	ser	5	птэ	ser	ser	ASII	10	Ala	ASII	1111	GIII	15	гуу	
129		7.00	T 110	Tlo	-	Tla	71.	uio	71 ~~~		Ala	Cox.	C1	Ф		Dwo	
	ser	Asp	тÀ2	20	116	тте	Ата	uis	25	GIY	Ala	ser	GTA		теп	PIO	
130 131	Clu	uio	mb ν		C1	C 0 x	т	חות		71.	Dha	71.	C1 5	30	71.7	7 an	
132	GIU	UTS	35	ьeu	Giu	ser	гу	40	пеа	Ala	Phe	Ala	45	GIII	нта	ASP	
133	Тиг	T 011		Cln	λαρ	Tou	7 l a		Th∽	T 1/10	Asp	C1.,		T 011	1/21	Wal	
134	ıyı	50	GIU	GIII	Asp	ьец	55	Met	IIII	ъу	ASP	60 60	Arg	ьeu	vaı	vaı	
135	Tlo		7) C.D.	u: c	Dho	Tou		C1.,	T 011	Th.	Asp		7.1 ~	T	T ***C	Dho	
136	65	птэ	Asp	UTS	rne	70	Asp	сту	ьeu	1111	75	vaı	Ата	гÃ2	гу	80	
137		uic	λνα	шic	λνα		λου	Cly	λνα	Tur	Tyr	Wal	Tlo	λen	Dho		
138	110	1113	Arg	1113	85	пуз	изр	GTA	Arg	90	1 y 1	vaı	116	ASP	95	1111	
139	Lou	Luc	Glu	Tlo		Sar	Len	Glu	Mot		Glu	Aen	Dho	Glu		Mot	
140	ьeu	пуз	GIU	100	GIII	Ser	ьeu	GIU	105	1111	GIU	ASII	rne	110	1111	Met	
141	70.1 -2	Mo+	Dha		Aen	Dro	Gln	Glu		Dro	Arg	Luc	Lau		Gln	Lau	
142	ALG	Hec	115	GIII	Азр	110	GIII	120	ALG	110	Arg	БуЗ	125	110	GIII	ьeu	
143	Cvs	Thr		T.A11	Gln	Thr	Thr		His	Asn	Ile	Tle		Glu	Cvs	Val	
144	Cys	130	OIU	пси	OIII	1111	135	110	111.5	тор	1+0	140	пси	Oiu	Cys	VUL	
145	Tur		Lvs	Gln	Gln	T.e.n		Ara	Ara	Glu	Val		Asn	Phe	Δla	Phe	
146	145	Cys	L) U	0111	0111	150	БСС	11119	711 g	OIU	155	- y -	пор	1110	7.1± u	160	
147		Asp	Len	Cvs	Tle		Tur	Ara	Asp	Glv	Asn	Pro	Tur	Ala	Val		
148	: 9	1101		010	165		- 1 -	9	1101	170			- 1 -		175	-1-	
149	Asp	Lvs	Cvs	Leu		Phe	Tvr	Ser	Lvs		Ser	G] 11	Tvr	Ara		Tvr	
150		-1-	0,10	180	-10		- 1		185				- 1 -	190		- 1 -	
151	Cvs	Tvr	Ser		Tvr	Glv	Thr	Thr		Glu	Gln	Gln	Tvr		Lvs	Pro	
152	-1-	- 1	195		- 1	1		200					205		-1-		
153	Leu	Cvs		Leu	Leu	Ile	Ara		Ile	Asn	Cys	Gln		Pro	Leu	Cvs	
154		210					215	-1-			-1-	220	-1-			-1-	
155	Pro		Glu	Lvs	Gln	Ara		Leu	Asp	Lvs	Lys		Ara	Phe	His	Asn	
156	225			-1-		230				-1-	235		5			240	
157		Ara	Glv	Ara	Trp		Glv	Ara	Cvs	Met	Ser	Cvs	Cvs	Ara	Ser		
158		5	1	9	245		1	5	-1-	250		-1-	-1-		255		
159	Ara	Thr	Ara	Ara		Thr	Gln	Leu	Thr		Gly	His	His	His		His	
160	5		5	260					265		1			270			
161	His													- • -			
	<210	> SEC	O ID	NO:	5												
	<2112																
	<212				-												
	<213				Artii	ficia	al Se	equer	nce								
				_				-									



DATE: 11/26/2002 PATENT APPLICATION: US/09/581,976B TIME: 09:26:10

Input Set : A:\seqlist.txt

169	<220	> FE	ATUR	Ε:														
170	<pre><223> OTHER INFORMATION: Chimaeric protein (protein D from Haemoplilus</pre>																	
171	r																	
172		vi	rus	type	16)													
174	<4002	> SE	QUEN	CE:	5													
1 7 5	atgo	gato	caa	gcag	ccat	tc a	tcaa	atat	g gc	gaat	accc	aaa	tgaa	atc	agac	aaaato	:	60
176																aaagca		120
177																gatggt		180
178					-	-			_	_	_	_		-	_	aaatto		240
179	_	_				_			_		_	_	_	-	_	gaaatt		300
180																		360
181																		
182																		
183																		
184														600				
185													660					
186														720				
187														780				
188																ttgcaa		840
189																gaggat		900
190	-			_			_	-	-		-		_					960
191														1020				
192														1080				
193														1116				
	<210					ca c	Cacci	acca	c ca	LLaa								1110
	<2112																	
	<212				1													
	<213				Arti.	fici	2) C	201121	200									
	<220				11. 6.1.	1101	ur o	cque	100									
					ያለው ተ	TON.	Chir	naer	ic n	rote	in (arati	sin 1) fr	om H:	emopl	i 1 111	2
202	12237								n fro		-	•			Om In	тепорт	. I I U	3
203				type		а по	U, T.	u5101		J111 111	uman	pap.	11101	ııa				
	<400>																	
206			-			His	Ser	Ser	Asn	Met	Δla	Asn	Thr	Gln	Met	Lvs		
207	1	тор	110	DCI	5	1113	DCI	UCI	71511	10	111.0	ASII	1111	OIII	15	цуз		
208		Asn	Lvs	Tle	_	Tle	Δla	His	Ara		Δla	Sar	Glv	Тиг	Leu	Pro		
209	Jei	АЗР	шуз	20	116	116	nra	1113	25	ОТУ	лта	261	СТУ	30	ьеu	110		
210	Glu	Hie	Thr		Glu	Sar	Luc	Δ 1 =		Δla	Dha	Λla	Gln		Ala	7 cn		
211	Olu	111.5	35	пец	Giu	261	шуз	40	пец	пта	1116	ліа	45	GIII	AIG	дэр		
212	Тиг	Lou		Cln	Λcn	T 011	Nlα		Thr	Tvc	Λcn	Clu		Ton	Val	Wal		
213	тут	50	Giu	GIII	ASP	пеп	55	net	1111	гуз	ASD		Arg	Leu	vaı	vaı		
213	Tlo		7.00	114.0	Dha	T 0		C1	T 0.11	mh	7	60	חות	T	T	DL -		
		птѕ	Asp	птѕ	Pne		ASP	GLŸ	ьeu	1111		vaı	ATa	гус	Lys			
215 216	65 Bro	u; ~	71	u:~	7. ~~	70	7.~~	C1	7~~	ጥ~	75 ∽	77 - 7	т1-	7	Dh.	80 Th~		
	PIO	HIS	Arg	HIS	_	гуѕ	ASP	GIY	Arg	_	ıyr	vaı	тте	Asp	Phe	Thr		
217	T ~··	T	C1	т1 -	85	C = ==	Tarr	C1	Mot	90 The	C1	7\	DI	C1	95	Mat		
218	ьeu	туѕ	GIU		GTU	ser	ьeu	GIU		ınr	GIU	Asn	rne		Thr	Met		
219	7.1	Met	DI	100	7\	D	C1-	C1	105	D	n	T	T	110	C1 -	T		
220 221	AIA	Mec		GIN	ASP	PFO	GTU	120	Arg	Pro	Arg	гÀг		rro	Gln	ьeu		
441			115					120					125					

DATE: 11/26/2002

TIME: 09:26:10



PATENT APPLICATION: US/09/581,976B

Input Set : A:\seqlist.txt
Output Set: N:\CRF4\11262002\I581976B.raw

222	Cvic	Th∽	Clu	T 011	Cln	Thr	Th.~	т1.	II i o	7 00	T10	Tlo	T 011	Clu	Cuc	17-1	
222	Cys		GIU	ьeu	GIII	TIII		тте	His	ASP	тте		ьеи	GIU	Cys	vai	
223	_	130		6 1	~ 1		135	_	_	0.3		140		D1		D)	
224		Cys	гàг	GIn	GIn		Leu	Arg	Arg	GLu		Tyr	Asp	Pne	Ата		
225	145	_				150	_				155	_	_			160	
226	Arg	Asp	Leu	Cys		Val	Tyr	Arg	Asp		Asn	Pro	Tyr	Ala		Cys	
227					165					170					175		
228	Asp	Lys	Cys	Leu	Lys	Phe	Tyr	Ser	Lys	Ile	Ser	Glu	Tyr		His	Tyr	
229				180					185					190			
230	Cys	Tyr	Ser	Leu	Tyr	Gly	Thr	Thr	Leu	Glu	Gln	Gln	Tyr	Asn	Lys	Pro	
231			195					200					205				
232	Leu	Cys	Asp	Leu	Leu	Ile	Arg	Cys	Ile	Asn	Cys	Gln	Lys	Pro	Leu	Cys	
233		210	_				215	-			-	220	-			_	
234	Pro	Glu	Glu	Lys	Gln	Arg	His	Leu	Asp	Lvs	Lvs	Gln	Arq	Phe	His	Asn	
235	225			-		230			-	4	235		,			240	
236		Ara	Glv	Ara	Trp		Glv	Ara	Cys	Met		Cvs	Cvs	Ara	Ser		
237		9		9	245		0 -1	9	o j o	250	001	010	0,10	9	255	001	
238	Ara	Thr	Ara	Ara		Thr	Gln	T.e.11	Met		Glv	Asp	Thr	Pro		I.eu	
239	7119	1111	1119	260	OIU	1111	OIII	ncu	265	1113	O ± y	пор	1111	270	1111	БСС	
240	Hic	Glu	ጥህን		Tan	Aen	LOU	Gln	Pro	Glu	Thr	Thr	Aen		Тиг	Cve	
241	1113	Gru	275	1100	neu	тэр	neu	280	110	GIU	1111	1111	285	пец	тут	Cys	
242	Ф	C1,,		T 011	λαη	7 cn	C02		Glu	Clu	C1,,	7.00		т1.	7 cn	C1	
243	тут	290	GIII	пеп	WOII	Asp	295	ser	Giu	GIU	Gru	300	GIU	TTE	нэр	GIY	
	Dro		C1.,	Cln	ת ות	C1,,		7 00	7 ~ ~	7.1 -	uic		7 on	Tlo	Wal	ሞኮ∽	
244 245	305	Ald	GIY	GIII	Ald	310	PLO	Asp	Arg	Ald	315	тÀт	ASII	тте	νат	320	
		C	C	T	C		C	m\	т	7)		C	777	C1	0		
246	Phe	Cys	Cys	гÀг		Asp	ser	Inr	Leu	_	ьeu	Cys	val	GIII		Thr	
247	112 -	77 - 7	70	т1.	325	m1	т	C1	7	330	τ	M = 4	C1	m1	335	, (1)	
248	HIS	vaı	Asp		Arg	Thr	ьeu	GIU	Asp	ьeu	Leu	мет	СТА		Leu	GTA	
249	T 3 .	** 7		340	~ 1	^	_	63	345	Б	m)	.	~ 1	350		** *	
250	TTE	vaı		Pro	тте	cys	Ser		Lys	Pro	Thr	ser		HIS	HIS	HIS	
251			355					360					365				
252	His	His	His														
253		370			_												
	<210																
	<211:				3										•		
	<212																
	<213				Arti:	ficia	al Se	equer	ıce								
	<220										_						
	<223														om Ha	aemoplilus	
262							ated	1 E7	fron	n Hur	nan p	papil	Lloma	3			
	<400																
266	atg	gatco	caa 🤄	gcago	ccatt	c at	caaa	atato	g gcg	gaata	accc	aaat	gaaa	atc a	agaca	aaatc	60
267																aaagca	120
268	ctt	gcgtt	itg d	cacaa	acag	gc to	gatta	attta	a gag	gcaag	gatt	tago	caato	gac t	aagg	gatggt	180
269	cgtt	tagt	igg t	tatt	cac	ga to	cactt	ttta	a gat	ggct	tga	ctga	atgti	gc g	gaaaa	aaattc	240
270	ccad	catco	gtc a	atcgt	aaaq	ga to	ggccg	gttac	tat	gtca	atcg	actt	taco	ctt a	aaaa	gaaatt	300
271	caaa	agtti	ag a	aato	gacad	ga aa	actt	tgaa	acc	catgo	gcca	tgca	atgga	aga t	cacac	cctaca	360
272																cagcaa	420
273																gaaccg	480
				-				-		•	•	•				- •	

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/581,976B

DATE: 11/26/2002 TIME: 09:26:11

Input Set : A:\seqlist.txt